Fault-tolerance

6.033 Lecture 14
Sam Madden
Road Map

• So far: Modularity & Operating Systems; Networked Systems

• Rest of this semester:
  • - how to keep running despite failures
  • - broaden class of failures to include malicious ones
Fault Tolerance Plan

- General introduction: today
  - Recovery/Replication
- Transactions: next 4 lectures
  - updating permanent data in the presence of concurrent actions and failures
- Replicated state machines: 2 more
  - Keep computing despite failures
Your computer restarted because of a problem. Press a key or wait a few seconds to continue starting up.

Votre ordinateur a redémarré en raison d’un problème. Pour poursuivre le redémarrage, appuyez sur une touche ou patientez quelques secondes.

El ordenador se ha reiniciado debido a un problema. Para continuar con el arranque, pulse cualquier tecla o espere unos segundos.

Ihr Computer wurde aufgrund eines Problems neu gestartet. Drücken Sie zum Fortfahren eine Taste oder warten Sie einige Sekunden.

問題が起きたためコンピュータを再起動しました。そのまま起動する場合は、いずれかのキーを押すか、数秒間そのままお待ちください。

电脑因出现问题而重新启动。请按一下按键，或等几秒钟以继续启动。
Windows

A fatal exception 0E has occurred at 0028:C00068F8 in PPT.EXE<01> 0000059F8. The current application will be terminated.

* Press any key to terminate the application.
* Press CTRL+ALT+DEL to restart your computer. You will lose any unsaved information in all applications.

Press any key to continue
Your PC ran into a problem and needs to restart. We're just collecting some error info, and then we'll restart for you. (0% complete)

If you'd like to know more, you can search online later for this error: HAL_INITIALIZATION_FAILED
San Francisco plane crash caused by pilot's inexperience with onboard computers

*Boeing 777's auto-throttle did not maintain speed as expected*

*By Aaron Souppouris on December 12, 2013 03:08 am  Email  @AaronIsSocial*
Router glitch cuts Net access

What started out as a router glitch at a small Internet service provider in Virginia today triggered a major outage in Internet access across the country, lasting more than two hours in some places.
Relative frequency of hardware replacement

<table>
<thead>
<tr>
<th>Component</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power supply</td>
<td>34.8</td>
</tr>
<tr>
<td>Memory</td>
<td>20.1</td>
</tr>
<tr>
<td><strong>Hard drive</strong></td>
<td>18.1</td>
</tr>
<tr>
<td>Case</td>
<td>11.4</td>
</tr>
<tr>
<td>Fan</td>
<td>8.0</td>
</tr>
<tr>
<td>CPU</td>
<td>2.0</td>
</tr>
<tr>
<td>SCSI Board</td>
<td>0.6</td>
</tr>
<tr>
<td>NIC Card</td>
<td>1.2</td>
</tr>
<tr>
<td>LV Power Board</td>
<td>0.6</td>
</tr>
<tr>
<td>CPU heatsink</td>
<td>0.6</td>
</tr>
</tbody>
</table>

10,000 machines

Pr(failure in 1 year) \(\sim 0.3\)

Schroeder and Gibson, FAST 2008
Availability in practice

- Carrier airlines (2002 FAA fact book)
  - 41 accidents, 6.7M departures
  - 99.9993% availability

- 911 Phone service (1993 NRIC report)
  - 29 minutes per line per year
  - 99.994%

- Standard phone service (various sources)
  - 53+ minutes per line per year
  - 99.99+% 

- End-to-end Internet Availability
  - 95% - 99.6%
**Barracuda® 7200.10**

Experience the industry’s proven flagship perpendicular 3.5-inch hard drive

**Key Advantages**
- First 3.5-inch drive to utilize capacity- and reliability-boosting perpendicular recording technology
- First drive to reach 750 GB—a full year ahead of competition—enabling new solutions for data-intensive applications.
- Industry’s most proven and established desktop hard drive available today—more than 16 million shipped to date.
- “One-stop shopping” with a broad range of capacity, cache and interface options for all your computing needs.
- Best-in-class environmental specifications and reliability features.
- Adaptive fly height offers consistent read/write performance from the beginning to the end of your computing workload.
- Clean Sweep automatically calibrates your drive.
- Directed Offline Scan runs diagnostics when storage access is not needed.
- RoHS-compliant design assures an environmentally conscious product.
- Enhanced G-Force Protection™ defends against handling damage.
- Seagate® SoftSonic™ motor enables whisper-quiet operation.

**Best-Fit Applications**
- Desktop and High-Performance PCs
  - Gamer PCs
  - Workstations
  - High-end PCs
  - Desktop RAID
  - Mainstream PCs
  - Point-of-sale devices/ATMs
  - USB/Firewire/sATA personal external storage

---

**Contact Start-Stops** | 50,000

**Nonrecoverable Read Errors per Bits Read** | 1 per $10^{14}$

**Mean Time Between Failures (MTBF, hours)** | 700,000

**Annualized Failure Rate (AFR)** | 0.34%
**Barracuda® ES.2**
High-capacity, business-critical Tier 2 enterprise drives

1 TB, 750 GB, 500 GB and 250 GB • 7200 RPM • SATA 3Gb/s, SATA 1.5Gb/s and SAS 3Gb/s

<table>
<thead>
<tr>
<th>Reliability/Data Integrity</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Time Between Failures (MTBF, hours)</td>
<td>1.2 million</td>
</tr>
<tr>
<td>Reliability Rating at Full 24x7 Operation (AFR)</td>
<td>0.73%</td>
</tr>
<tr>
<td>Nonrecoverable Read Errors per Bits Read</td>
<td>1 sector per 10E15</td>
</tr>
<tr>
<td>Error Control/Correction (ECC)</td>
<td>10 bit</td>
</tr>
<tr>
<td>Interface Ports</td>
<td></td>
</tr>
<tr>
<td>SATA</td>
<td>Single</td>
</tr>
<tr>
<td>SAS</td>
<td>Dual</td>
</tr>
</tbody>
</table>
Disk failure conditional probability distribution

- Infant mortality

- Burn out

- Stable failure period

- Expected operating lifetime

- 1 / (reported MTTF)

Bathtub curve

- (~5 years)
Disk Age vs. $\Pr(\geq 1 \text{ Reported Read Failure})$

$\Pr(\text{Reported Read Failure})$

Disk Age (Months)

Bairavasundaram et al., SIGMETRICS 2007
Fail-fast disk

```c
failfast_get (data, sector) {
    get (s, sector);
    if (checksum(s.data) = s.cksum) {
        data ← s.data;
        return OK;
    } else {
        return BAD;
    }
}
```
Careful disk

careful_get (data, sector) {

    r ← 0;

    while (r < 10) {

        r ← failfast_get (data, sector);

        if (r = OK) return OK;

        r++;
    }

    return BAD;
}
Replicated Disks

write (sector, data):
    write(disk1, sector, data)
    write(disk2, sector, data)

read (data, sector):
    data = careful_get(disk1, sector)
    if error
        data = careful_get(disk2, sector)
        if error
            return error
    return data
## Technical specifications

<table>
<thead>
<tr>
<th>Specification</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processors</td>
<td>2–16 per node&lt;br&gt;Intel Itanium processor 9100 series processors, 1.6 GHz single core processors</td>
</tr>
<tr>
<td>Cache</td>
<td>12 MB L3</td>
</tr>
<tr>
<td>RAM standard/maximum</td>
<td>Minimum: 4 GB&lt;br&gt;Maximum: 16 GB (32 GB²)</td>
</tr>
<tr>
<td>RAM type/speed</td>
<td>PC2100 ECC registered DDR266A/B</td>
</tr>
<tr>
<td>ServerNet I/O</td>
<td>Minimum: 10&lt;br&gt;Maximum: 60</td>
</tr>
<tr>
<td>I/O adapters supported</td>
<td>Fibre Channel, Gigabit Ethernet</td>
</tr>
<tr>
<td>Fibre Channel disk modules</td>
<td>14 disks per module</td>
</tr>
<tr>
<td>Disk drives supported</td>
<td>146 GB and 300 GB 15K RPM Fibre Channel internal hard disk drive drives&lt;br&gt;HP Disk Array family (e.g., XP24000, XP20000, XP12000, and XP10000 disk arrays)</td>
</tr>
<tr>
<td>Standard features</td>
<td>N + 1 power supplies&lt;br&gt;N + 1 fans</td>
</tr>
</tbody>
</table>

² Although 32 GB possible, the Integrity NonStop NS16200 Server support availability is up to 16 GB.
How about an error in software?

- Big problem!
- Software for fault tolerant systems must be written with great care
  - Stringent development practices
  - Well-defined stable specification
  - Modeling, simulation, verification, etc.
  - N-version programming is tricky
- Will also be a problem for secure software
- Good design: small fraction is critical